EXHIBIT I

Target amplification for DNA analysis by the polymerase chain reaction

K.B. MULLIS [1]

SUMMARY

The polymarase chain raction (PCR) has becoma a standard laboratory technique. An anzymatic reaction, as simple to perform as it is satisfying to contamplete, the PCR solvas two of the mora univarsal problems in the chemistry of naturel nucleic acids. It allowa for tha physical saparetion any particular sequence of interast from its context ; and than provides for an in vitro amplification of this aequanca which is virtually without limit. The surprising robustness of PCR derivas from its fortuitoua combinetion of thraa phenomena, each of which is intrinsically powarful. The first of thaae is tha impreasive ability of almost all oligodeoxynucleotides to bind tightly and specifically to their complementary nucleic acid aequences, discriminating easily between hundreds of thousands of sites. The second familiar phenomenon is illustrated by the notion that the probebility for tha occurance of a compound action is the product of the individual probabilitias for the occurence of aach of its componants. Tha third phanomanon embodied in tha polymarasa chain reaction relatas to the branching structure of its propagation and tha inherent robustnass attachad to such a form. Considaration of tha abova leads to certain genaralities regarding the relativa utility of various protocols for carrying out the PCR. Specific conditions of time, tamperatures, concentretions, etc. will be described, as wall as aample preparation and analytical methods.

INDEX TERMS : DNA polymerase - DNA probes

RÉSUMÉ

La polymerase chain reaction : technique d'amplification pour l'analyse de l'ADN. – La Polymerase Chain Reaction (PCR) est devenue une technique largement utilisée en laboratoire. Cette reaction enzymatique, aussi simple a mettre en œuvre que satisfaisante, permet de resoudre deux des plus importants problèmes de la chimie des acides nucleiques naturels. Elle permet la separation physique de la sequence qui nous intéresse de son contexte, et l'amplification de cette sequence sans limite théorique. La surprenante robustesse de la PCR est le resultat de la combinaison fortuite de trois phenomènes, possedant chacun une puissance intrinseque. Le premier de ces phenomenes est l'impressionnante capacite de la majorite des oligodeoxynucleotides de se lier specifiquement et fortement a leurs sequences d'acides nucleiques complementaires, assurant une discrimination aisee parmi des centaines de milliers de sites. Le second phenomene bien connu est illustre par la notion que la probabilite qu'un evenement complexe se realise est le produit des probabilites de chaque element qui compose cet evenement. Le troisième phenomène, specifique a la polymerase chain reaction, est relatif au mecanisme d'extension par derivation et a la robustesse inherente liee a cette forme. Les considerations ci-dessus nous conduisent a envisager un certain nombre de generalites concernant l'interêt relatif des divers protocoles de mise en œuvre de la PCR. Les conditions specifiques de temps, de temperatures, de concentrations, etc. sont decrites, ainsi que la preparation de l'echantillon et les methodes analytiques.

MOTS CLES : ADN polymerase - Sonde.

The polymerase chain reaction [1-3] is suddenly a standard laboratory technique. An enzymatic reaction, as simple to perform as it is satisfying to contemplate, the PCR solves two of the more universal problems in the chemistry of natural nucleic acids. It allows for the physical separation of any particular sequence of interest from its context ; and then provides for an in vitro amplification of this sequence, virtually without limit. How is it that a simple arrangement of two oligodeoxynucleotides, four deoxynucleoside triphosphates and a DNA polymerase can produce such a result ?

The surprising robustness of PCR, the fact that it is difficult to find situations in which it will not work, derives from its fortuitous combination of three familiar phenomena, each of which is intrinsically powerful.

The first of these is the impressive ability of almost all oligodeoxynucleotides under a wide range of conditions to bind tightly and specifically to their complementary nucleic acid sequences. These short single-stranded DNA molecules are able to sort through, as it were, hundreds of thousands of likely binding sites, discriminating easily between a perfect fit and an almost perfect fit ; binding tightly to the perfect complement ; and at practical concentrations and temperatures accomplishing this in milliseconds. One in a hundred-thousand is a respectable shot. If the human genome contained only a few hundred-thousand base pairs, the selectivity of an oligodeoxynucleotide alone would enable most conceivable discriminations to be made readily. The human genome, however, consists of about 3.3×10^9 base pairs, and as had been pretty well demonstrated [4] prior to the emergence of PCR, something in addition to oligodeoxynucleotide specificity is necessary in order to make the specific sampling of human DNA sequence variation possible. Theoretically, the information contained in a seventeen-base-long oligodeoxynucleotide is sufficient to define a precise location on a random sequence the size of the human genome ($4^{17} = 1.72 \times 10^{10}$ is greater than 6.6×10^9 which is the number of nucleotides available for hybridization when the human genome is denatured). However absolute specificity is not obtained in practice even at high temperatures, and there are good reasons why not. Most of them have to do with the fact that if something else can happen, it does. Using longer oligodeoxynucleotides and/or involved and/or secret proto cols doesn't markedly improve the situation. Thus, oligonucleotides provide a good starting point for procedures which must address specific sequences, but they are not in themselves sufficient.

[1] Biotechnology Consultant, 6767 Neptune Place, Apt. 4, La Jolla, California, USA.
Reprints : K.B. Mullis, address above.

Conference presented at the First Congress on Advanced Biotechnology. Paris, December, 5-8, 1989.

The second familiar phenomenon is illustrated by the notion that the probability for the occurence of a compound action is the product of the individual probabilities for the occurrence of each of its components. What is this got to do with the PCR ? Everything. This is also the critical feature of « sandwich » format immunochemical assays. In both cases « correct » binding events which are very probable, are unavoidably associated with a very much larger number of « incorrect » binding events. The « incorrect » events are individually improbable but the whole hoard of them collectively constitute an unacceptably high background. The polymerase chain reaction like the « sandwich » assays succeeds by strictly requiring that a final signal can only be generated by the coordinated action of two binding events. In the case of the PCR, a fragment is only amplified if two oligodeocynucleotides are bound to the single-stranded DNA templates in such a way that the polymerase extension product of the one contains a binding site for the other. The probability for this in the presence of the intended DNA target is the product of the probabilities of the individual oligodeoxynucleotides finding their binding sites on the target DNA strands. Under appropriate conditions of concentration, time and temperature, the individual binding probabilities are close to 1.0, and therefore, so is their product. Therefore, during every cycle of PCR, there is a high probability that a new copy of the target is produced from each existing copy. However, for any particular non-target stretch of DNA to be amplified, two probabilities far less than one (maybe 10^{-5}, that being the approximate probability that a single oligodeoxynucleotide will bind to a random DNA sequence) have to be multiplied, and the 4 probability of the compound action is therefore very low (maybe 10^{-10}). The statistical treatment there is incomplete but should be suggestive of the way the PCR improves remarkably on the degree of specificity that can ultimately be obtained from the binding of oligodeoxynucleotides to targets on complex DNA.

The third familiar phenomenon embodied in the polymerase chain reaction relates to the properties of a chain reaction itself. The PCR is a sort of deoxyribonuclear bomb. Exponential growth is obtained by repeated copying of templates into new templates. This undoubtedly imbues the PCR with much of its aforementioned robustness. PCR amplifications are in this way also similar to genes pools. Both are harder to terminate with a single blow because they are carried in a population which is continually dividing, wherein the malfunction of an individual is not fatal for the pool. A linear process can be cut off at a single source, not so with a chain reaction. The process of division provides more and more growing points. Exponential growth of any kind is difficult to step on, but exponential growth, which is a result of a chain reaction, in which products become reactants, is a particularly viable beast. Exponential growth is also important from a practical point of view with regard to the time required for a given level of amplification. But this is only a matter of convenience. If we had to wait, we could.

Amplification is the most striking result of PCR. It has allowed us to look at single molecules and to look in biological samples for very rare occurences. However, amplification aside, the ability to easily prepare pure DNA sequences is starting to have an impact on the way molecular biology is done and on the type of experiments that a molecular biologist will consider. Relatively pure fragments of DNA, the ends of which are not dependent on the fortuitous presence of restriction enzyme sites are now accessible. And these can be modified easily. The fact that the 5' end of the oligodeoxynucleotides is not critical to the PCR reaction itself allows the experimenter the option to insert functional DNA sequences, or non-nucleic-acid substituents onto the ends of the targeted fragment. Thus, restriction enzyme sites, RNA polymerase promoter sequences, organic chemical handlees for capturing or detecting a sequence, etc., can be easily appended as the fragment of interest is being isolated. The problem of getting flanking sequences by a PCR mechanism is one that a lot of people have thought about for a long time and finally this year it was solved in a particularly nice way and apparently in more than one lab simultaneously [5, 6].

The trick is to circularize the target DNA and turn the primers for a known region around backwards thereby amplifying the flanks. In addition to other uses this clearly will be a real boon for sequencing the human genome.

Some practical comments

The precise conditions under which PCR always works best are not known and it seems reasonable to assume that no such standard conditions will be discovered. After all, people use PCR for very different purposes. The 1 000-fold amplification of a plasmid-born fragment starting from 100 femtomoles is different from the 10^{11}-fold amplification of several molecules up to 100 femtomoles [7]. In addition amplifiying a 58-bp fragment is not the same as amplifying a 10-kb fragment [8]. There are some things however, which are generally true and which may be useful to keep in mind when setting up a particular reaction.

Enzyme and primer concentration

One unit of the polymerase from Thermus aquaticus is about 50 femtomoles of the protein [9]. This compares with about 500 femtomoles per unit for Klenow. If you are using 3 units of Taq in a 100 microliter reaction then the enzyme is 1.5 nM. If you are amplifying a fragment up to the 10 nM level, which is where a 5-10 microliter aliquot becomes readily visible on an ethidium stained gel (depending on the number of bp in the fragment) ; then you should consider the fact that enzyme will become limiting at some point. At fragment concentrations greater than 1.5 nM some molecules of the enzyme are being asked to extend two fragments per cycle.

The success of the PCR reaction depends on the kinetic advantage that high concentrations of primers have over relatively low concentrations of product strands which at equilibrium would rehybridize

with each other and displace the primers. This advantage is lost if sufficient enzyme is not available during the moments immediately after the primers have hybridized. Likewise, the primer concentrations must be high enough to begin with, so that the primers do indeed get there first. Primer concentrations less than 100-fold greater than expected product concentrations would be asking a lot more than oligonucleotides can probably provide. The relative importance of these parameters is hard to dissect from the available data, but the most impressive amplifications which I've seen, in terms of very high gain amplifications [7] is amplification of one or several fragments to high levels [3-10] ; or amplification of very large fragments [8] which have all used large amounts of enzyme (5-15 units per 100 μl), and primers between one and ten micromolar. For amplifications requiring many cycles, enzyme has been added more than one.

Reaction time

This can be very short for fragments in the 50-200 bp range. I suggest trying no time at all first. Cycle back and forth as fast as your cycler will allow between 65 and 98 degrees. Compare with dropping to 60 and/or stopping at 75 for 2 or more minutes. For long fragments, you may have to wait 10 or 20 minutes. The temperature optimum for the enzyme itself on an M13 template is around 74.

From the molecular weight of 94,000 and specific activity of ca. 250 000 units/mg [11] one can calculate a turnover number or minimum extension rate at 74 degrees of 131 nt/sec. The extension rate determined from labelled primers [9] is 60 - 120 nt/sec at 80 degrees, but the precise relevance of these numbers to PCR reactions in not exactly clear. It has been observed in several labs that longer fragments can require 10-20 minutes per extension cycle to achieve optimum yields, which still are lower than short fragment yields, which still are lower than short fragment yields. If you are going to be amplifying the same fragment over and over again, you should do some experiments to determine the best times. Theoretically, the shortest possible reaction times which are required to achieve a complete duplication of your intended product, will be optimal. Anything further will work to your disadvantage.

Magnesium, dNTPs, and melting temperature

Every double stranded DNA does not denature at the same temperature, not even close. It is not known wether complete denaturation is necessary for amplification, but it is known that a number of amplifications have failed because of insufficient heating. For the time being we are stuck with an enzyme which cannot be boiled, which is a shame. We are using this particular enzyme, not because it is the best possible enzyme for the job, but because I somewhat arbitrarily picked it from the four or five thermophilic enzymes which were described in the literature by 1984, and after it was purified in David Geldand's lab [11], and we found that it worked okay for PCR, nobody saw fit to purify any other ones. Someone should, but in the meantime we have to deal with an enzyme that begins to seriously fall apart around 95 degrees, and in some cases that causes trouble. If your fragment is particularly hard to melt, a higher temperature, lower salts, or a little dimethylsulfoxide, might all help raise the yield. Higher temperatures for extended times are out ; and the enzyme requires (according to Bill Sutherland of Eastman Kodak) at least one millimolar free magnesium, meaning concentrations of magnesium one millimolar greater than the total concentration of all the dNTPs (which have a taste for magnesium themselves). So you're between a rock and a hard place again. Some people lower the magnesium anyhow and consequently must also lower the dNTPs. By running under conditions that are near marginal for magnesium, they are risking variable results if samples are introduced which contain various levels of magnesium. Sodium and/or potassium are more or less ornamental components of some recipes that are not necessary for the reaction, and can be left out entirely without hurting anything. Dropping the monovalent ion concentration by 50 mM is equivalent to dropping the magnesium concentration by 5 mM in terms of destabilizing double stranded DNA. And 10% DMSO has been added from time to time without disastrous side effects. None of these parameters have been studied sufficiently by anybody to raise the level of this discussion higher than just the local gossip on the PCR circuit. The reason for this is perhaps that most of the time no such efforts are necessary. The fragments melt easily below 95 degrees and the reactions work fine under any reasonable conditions.

Out of a natural laziness I always start with the easiest possible protocol and work from there. Or better yet, suggest that someone else start from there, and come back in a month to see how things worked out. The easiest possible protocol is one where the numbers are all round and even powers often, so I start from :

100 mM Tris-HCl at pH 9.0*
10 mM $MgCl_2$
1-10 uM oligodeoxynucleotides
1 mM dNTPs
10 units of Taq/100 μl reaction

Remembering that long DNA molecules are more resistant to heat than relatively small PCR fragments, thoroughly denature the DNA sample prior to adding the enzyme by immersing the reaction tube in boiling water for 2 minutes. There are several reports from different sources indicating that more is involved here than just denaturing the target. My suspicion is that in certain cases at least it is important not to have very long pieces of DNA in the reaction, the rationale being that a mispriming event on a 100 kb strand is more serious than one on a 1 kb strand. Boiling water will certainly decrease the average size of native DNA. Either add the enzyme before the solution cools down or bring the machine up to the lowest temperature you will be using (65 degrees) and then add it. Don't let the temperature

*Measured at room temperature whatever it happens to be.

drop below 65 while the enzyme is there until the reaction is over. These precautions will not have a noticeable effect until you begin to amplify very small numbers of molecules for many cycles at which point they may be critical. Start cycling back and forth between 65 and 98 degrees as fast as you can. If possible begin with 10 femtomoles of a cloned target. Do ten cycles, put ten microliters from the last cycle on an agarose gel with 0.5 μg/ml ethidium bromide in it, and look for a band after a few minutes. If it works then try more cycles from less sample. If it doesn't, vary the parameters until it does. Get it right at this level before you go on to something more demanding, and don't do anything you don't have to ; because every step you add to your protocol is likely to ruin it.

Be aware of the fact that the polymerase from Thermus aquaticus is very capable of working at low temperatures and the worst thing you can do to a PCR reaction is to allow extension of primers to occur at a low temperature. By doing so you generate a large number of incorrect extension products with the promiscuous oligomer at one end. Some of them, especially the long ones, will provide a site for a second primer to bind, and then you have a fragment which can be amplified. But it isn't the one your grant was written about. Enough of these voluntary fragments can choke a PCR reaction to death at the level of the finite supply of polymerase, before the intended target has been amplified to the level required by your detection system.

The recipes employed by the two groups who have presented what I thought were the most impressive PCR results to date [7-10], did not employ round numbers. They both use 67 mM Tris at pH 8.8, 16.7 mM ammonium sulfate, 6.7 mM magnesium chloride, BSA, and other less cryptic things which can be found in their papers. I tried adding 20 mM ammonium sulfate to my even powers of ten buffer, referred to on the privious page, and for reasons that I don't understand yet, it noticeably improves the signal to noise ratio in two different systems where 300 copies of a sequence are being amplified in the presence of 1 μg total DNA. I don't like doing things I don't understand, but ammonium sulfate is cheap.

REFERENCES

1. Mullis K.B. – US patent 4, 683, 202 Process for amplifying Nucleic acids sequences. Filed 10-25-85, Issued 7-28-87.
2. Mullis K.B., Faloona F., Scharf S., Saiki R., Horn G., Erlich H. – Specific enzymatic amplification of DNA in vitro : the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, 1986, 51, 263.
3. Mullis K.B., Faloona F. – Specific synthesis of DNA in vitro via a polymerase-catalysed chain reaction. *Meth. Enzymol.*, 1987, 155, 335.
4. Wallace R.B., Petz L.D., Yam P.Y. – Application of synthetic DNA probes to the analysis of DNA sequence variants in man. *Cold Spring Harbor Symposium on Quantitative Biology*, 1986, 51, 257.
5. Triglia T., Gregory Peterson M., Kemp D.J. – A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nucl. Acids Res.*, 1988, 16, 8186.
6. Ochman H., Gerber A.S., Hartl D.J. – Genetic applications of an inverse polymerase chain reaction. *Genetics*, 1988, 120, 621.
7. Kim H., Smithies O. – Recombinant fragment assay for gene targetting based on polymerase chain reaction. *Nucl. Acids Res.*, 1988, 16, 8887.
8. Jeffreys A.J., Wilson V., Newmann R., Keyte J. – Amplification of human minisatellites by the polymerase chain reaction : towards DNA fingerprinting of single cells. *Nucleic Acid Res.*, 1988, 16, 10953.
9. Innis M.A., Gelfand D.H., Myambo K.B., Brow M.D. – DNA sequencing with thermus aquaticus DNA polymerase and direct sequencing of polymerase chain amplified DNA. *Proc. Natl. Acad. Sci.*, 1988, 85, 9436.
10. Chamberlain J.S., Gibbs R.A., Ranier J.E., Nguyen P.N., Caskey C.T. – Deletion screening of the Duchenne Muscular Dystrophy locus via multiplex DNA amplification. *Nucl. Acid. Res.*, 1988, 16, t1141.
11. Gelfand D.H., Stoffel S. (in preparation)

48e année - 1990 N° 8
MENSUEL

ABCLAI 48, (8) 529 à 616 (1990)
ISSN 0003-3898

abc

Dans ce numéro :